

We claim:

1. A system comprising:

client computers having one or more data records, the client computers in communication with a network, the client computers configured to field-level normalize and one-way encrypt one or more fields of the one or more data records to provide one or more de-identified records; and

a server computer in communication with the network to receive the one or more de-identified records and in communication with a database, the database including one or more master records, the server computer configured to compare the one or more de-identified records with the one or more master records and to determine which records of the one or more de-identified records and the one or more master records are to be linked.

2. The system of claim 1 wherein the database is partially described by a table of master records.

3. The system of claim 2 wherein the table is for comparing the one or more de-identified records are compared with the one or more master records.

4. A method for de-identification of at least one record by a programmed client computer, comprising:

obtaining the at least one record, the at least one record having data fields;

normalizing at least a portion of the data fields; and

one-way hashing the at least a portion of the data fields to provide a de-identified record.

5. The method of claim 4 further comprising:

two-way encrypting the de-identified record;

compressing the de-identified record; and

transmitting the de-identified record.

6. The method of claim 5 further comprising encoding the data fields after normalization.

7. A method for de-identification of records by and at a programmed client computer, comprising:

providing records to the programmed client computer;

locating personal identification data fields in each of the records;

parsing the personal identification data fields;

formatting the personal identification data fields;

selecting at least a portion of the personal identification data fields formatted;

deleting any of the personal identification data fields not selected; and

one-way encrypting the personal identification data fields selected.

8. The method of claim 7 further comprising:

obtaining a mapping file; and

locating personal identification data fields in each of the records using the mapping file.

9. The method of claim 7 further comprising:

determining if the personal identification data fields selected are to be encoded; and

encoding the personal identification data fields to be encoded.

10. The method of claim 9 further comprising concatenating the personal identification data fields encoded with a seed value to provide seed value identifiers.

11. The method of claim 9 wherein the personal identification data fields are not concatenated with a seed value prior to the one-way encrypting.

12. The method of claim 7 wherein the one-way encrypting step comprises:
one-way encrypting with a first encryption algorithm the personal identification data fields selected to provide a first encryption result for each of the personal identification data fields selected; and

one-way encrypting with a second encryption algorithm the personal identification data fields selected to provide a second encryption result for each of the personal identification data fields selected.

13. The method of claim 12 wherein the one-way encrypting step comprises:

concatenating at least a portion of each of the first encryption result and the second encryption result for each of the personal identification data fields to respectively provide binary string identifiers for the personal identification data fields; and

converting the binary strings to alphanumeric strings to provide match codes.

14. A method for de-identification of records by a programmed client computer, comprising:

monitoring a file directory;

detecting presence of a new file in the file directory;

obtaining a mapping file for the new file;

locating personal identification data fields in records in the new file using the mapping file;

parsing the personal identification data fields;

formatting the personal identification data fields;

selecting at least a portion of the personal identification data fields formatted;

deleting any of the personal identification data fields not selected;

determining if the personal identification data fields selected are to be encoded;

encoding the personal identification data fields to be encoded;

concatenating the personal identification data fields encoded with a seed value to provide seed value identifiers;

first one-way encrypting the seed value identifiers with a first encryption algorithm;

second one-way encrypting the seed value identifiers with a second encryption algorithm;

concatenating at least a portion of each one-way encryption result from the first one-way encrypting and the second one-way encrypting corresponding to the seed value identifiers to respectively provide binary strings for each of the seed value identifiers; and

converting the binary strings to alphanumeric strings to provide match codes;

wherein de-identified records comprising the match codes are created at the programmed client computer prior to transmission to a server computer.

15. A method for linkage of de-identified records, comprising:

obtaining client de-identified records, the client de-identified records comprising field-level one-way hashed match codes;

providing a database of master de-identified records, the master de-identified records comprising field-level one-way hashed match codes;

comparing the match codes of the client de-identified records and the master de-identified records;

creating an initial match group and an initial no match group from the comparing of the match codes;

calculating individual weights for each comparison of match codes;

calculating a total match score from the individual weights;

calculating an upper threshold and a lower threshold;

placing each of the client de-identified records according to the total match score for each into one of a probable match group, a probable no match group and a statistical no match group;

repeating the calculating steps until change in record volume is within tolerance; and

linking at least a portion of the client de-identified records with the master de-identified records.

16. The method of claim 15 wherein the calculating steps are repeated using tabulations of the probable no match group and the statistical no match group.

17. The method of claim 16 further comprising adding the client de-identified records not linked with the master de-identified records to the database as new master de-identified records.

18. The method of claim 15 wherein the client de-identified records are from a plurality of data sources.

19. The method of claim 15 wherein the client de-identified records are created at a later date than creation of the master de-identified records.

20. The method of claim 15 wherein the linking is longitudinal linking.

21. A method for linkage of de-identified records, comprising:
obtaining client de-identified records, the client de-identified records comprising field-level one-way hashed match codes;
providing a database of master de-identified records, the master de-identified records comprising field-level one-way hashed match codes;
comparing the match codes of the client de-identified records and the master de-identified records;
deterministically creating an initial match group and an initial no match group from the comparing of the match codes;
calculating individual weights for each comparison of match codes of the client de-identified records and the master de-identified records;
calculating a total match score from the individual weights for each comparison of the client de-identified records and the master de-identified records;

calculating an upper threshold and a lower threshold, the upper threshold and the lower threshold in combination used to delineate a probable match group region, a probable no match group region and a statistical no match group region;

placing each of the client de-identified records according to the total match score for each into one of the probable match group region, the probable no match group region and the statistical no match group region to provide at least in part a probable match group and a statistical no match group;

comparing at least one of the probable match group or the statistical no match group to the initial match group or the initial no match group, respectively, for change in volume of records;

if the change in volume of records is not within a determined percentage, using the probable match group and the statistical no match group to calculate the individual weights; and

if the change in volume of records is within the determined percentage,

longitudinally linking the client de-identified records to the master de-identified records by appending record identifiers of the master de-identified records to the client de-identified records;

adding the client de-identified records not linked to the master de-identified records to the database as master records; and

appending the record identifiers to the client de-identified records added.

22. A signal-bearing medium containing a program which, when executed by a processor, causes execution of a method comprising:

obtaining at least one record, the record having data fields;

normalizing at least a portion of the data fields; and

one-way hashing the at least a portion of the data fields to provide a de-identified record.

23. A signal-bearing medium containing a program which, when executed by a programmed client computer, causes execution of a method comprising:

providing records to the programmed client computer;
locating personal identification data fields in each of the records;
parsing the personal identification data fields;
formatting the personal identification data fields;
selecting at least a portion of the personal identification data fields
formatted;
deleting any of the personal identification data fields not selected; and
one-way encrypting the personal identification data fields selected.

24. A signal-bearing medium containing a program which, when executed by a programmed client computer, causes execution of a method comprising:
monitoring a file directory;
detecting presence of a new file in the file directory;
obtaining a mapping file for the new file;
locating personal identification data fields in records in the new file
using the mapping file;
parsing the personal identification data fields;
formatting the personal identification data fields;
selecting at least a portion of the personal identification data fields
formatted;
deleting any of the personal identification data fields not selected;
determining if the personal identification data fields selected are to be
encoded;
encoding the personal identification data fields to be encoded;
concatenating the personal identification data fields encoded with a
seed value to provide seed value identifiers;
first one-way encrypting the seed value identifiers with a first encryption
algorithm;
second one-way encrypting the seed value identifiers with a second
encryption algorithm;
concatenating at least a portion of each one-way encryption result from
the first one-way encrypting and the second one-way encrypting

corresponding to the seed value identifiers to respectively provide binary strings for each of the seed value identifiers; and

converting the binary strings to alphanumeric strings to provide match codes;

wherein de-identified records comprising the match codes are created at the programmed client computer prior to transmission to a server computer.

25. The method of claim 24 wherein the programmed client computer comprises a mapper program, a parser program, a formatting program and an encoding program.

26. A signal-bearing medium containing a program which, when executed by a programmed server computer, causes execution of a method comprising:

obtaining client de-identified records, the client de-identified records comprising field-level one-way hashed match codes;

accessing a database of master de-identified records comprising field-level one-way hashed match codes;

comparing the match codes of the client de-identified records and the master de-identified records;

creating an initial match group and an initial no match group from the comparing of the match codes;

calculating individual weights for each comparison of match codes;

calculating a total match score from the individual weights;

calculating an upper threshold and a lower threshold;

placing each of the client de-identified records according to the total match score for each into one of a probable match group, a probable no match group and a statistical no match group;

repeating the calculating steps if change in record volume is not within tolerance; and

linking at least a portion of the client de-identified records with the master de-identified records.

27. The method of claim 26 wherein the calculating steps are repeated using tabulations of the probable no match group and the statistical no match group.
28. The method of claim 27 further comprising adding the client de-identified records not linked with the master de-identified records to the database as new master de-identified records.
29. The method of claim 26 wherein the client de-identified records are from a plurality of data sources.
30. The method of claim 26 wherein the client de-identified records are created at a later date than creation of the master de-identified records.
31. The method of claim 26 wherein the linking is longitudinal linking.
32. A system comprising:
a data warehouse, the data warehouse comprising at least one database including master de-identified records and de-identified longitudinally linked records to at least a portion of the master de-identified records;
at least one server computer in communication with the data warehouse;
at least one customer computer;
a network configured to place the at least one server computer in communication with the at least one customer computer for transmitting at least a portion of the at least one database to the at least one customer computer to populate a data mart database.
33. The system of claim 32 wherein the at least one server computer or the at least one customer computer comprises an application configured to provide a data product from the de-identified longitudinally linked records.

34. A method for linkage of de-identified records, comprising:
obtaining client de-identified records, the client de-identified records comprising field-level one-way hashed match codes;
providing a database of master de-identified records, the master de-identified records comprising field-level one-way hashed match codes;
comparing the match codes of the client de-identified records and the master de-identified records; and
linking at least a portion of the client de-identified records with the master de-identified records using comparison of the match codes.
35. The method of claim 34 further comprising assigning identification codes to the master de-identified records.
36. The method of claim 35 further comprising appending the identification codes of the master de-identified records to the client de-identified records.
37. A method for transforming personal identifying information to facilitate protection of privacy interests while allowing use of non-personally identifying information, comprising:
receiving data on an individual including personally identifying information, de-identifying the data at a client computer including field-level one-way encryption, transmitting the de-identified data to a server computer for record linkage, and using match codes created for the data at the client computer to link records at the server computer.